Article

Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning

Matheus Henrique Dal Molin Ribeiro [1,2], **Stéfano Frizzo Stefenon** [3],
José Donizetti de Lima [1,4,*], **Ademir Nied** [3], **Viviana Cocco Mariani** [5,6]
and Leandro dos Santos Coelho [2,5]

1   Department of Mathematics (DAMAT), Federal Technological University of Parana (UTFPR),
    Pato Branco (PR) 85503-390, Brazil; mribeiro@utfpr.edu.br
2   Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic
    University of Parana (PUCPR), Curitiba (PR) 80215-901, Brazil; leandro.coelho@pucpr.br
3   Electrical Engineering Graduate Program, Department of Electrical Engineering,
    Santa Catarina State University (UDESC), Joinvile (SC) 80215-901, Brazil;
    stefano.stefenon@udesc.br (S.F.S.); ademir.nied@udesc.br (A.N.)
4   Industrial and Systems Engineering Graduate Program (PPGEPS),
    Federal Technological University of Parana (UTFPR), Pato Branco (PR) 85503-390, Brazil
5   Department of Electrical Engineering, Federal University of Parana (UFPR),
    Curitiba (PR) 80060-000, Brazil; viviana.mariani@pucpr.br
6   Department of Mechanical Engineering, Pontifical Catholic University of Parana (PUCPR),
    Curitiba (PR) 80215-901, Brazil
*   Correspondence: donizetti@utfpr.edu.br

check for updates

Abstract: Electricity price forecasting plays a vital role in the financial markets. This paper proposes a self-adaptive, decomposed, heterogeneous, and ensemble learning model for short-term electricity price forecasting one, two, and three-months-ahead in the Brazilian market. Exogenous variables, such as supply, lagged prices and demand are considered as inputs signals of the forecasting model. Firstly, the coyote optimization algorithm is adopted to tune the hyperparameters of complementary ensemble empirical mode decomposition in the pre-processing phase. Next, three machine learning models, including extreme learning machine, gradient boosting machine, and support vector regression models, as well as Gaussian process, are designed with the intent of handling the components obtained through the signal decomposition approach with focus on time series forecasting. The individual forecasting models are directly integrated in order to obtain the final forecasting prices one to three-months-ahead. In this case, a grid of forecasting models is obtained. The best forecasting model is the one that has better generalization out-of-sample. The empirical results show the efficiency of the proposed model. Additionally, it can achieve forecasting errors lower than 4.2% in terms of symmetric mean absolute percentage error. The ranking of importance of the variables, from the smallest to the largest is, lagged prices, demand, and supply. This paper provided useful insights for multi-step-ahead forecasting in the electrical market, once the proposed model can enhance forecasting accuracy and stability.

Keywords: complementary ensemble empirical mode decomposition; electricity price forecasting; ensemble learning models; exogenous variables; short-term forecasting

1. Introduction

The electricity power market is an important research topic, which has been receiving much attention from research over the last years [1–3]. The high interest in this field is due to the forecasting

prices allowing for managers (short, medium, or long-term horizon) to use the forecasting information to adjust their finances, as well as develop strategic planning to support the decision-making system. The development of high accuracy forecasting models is hard, due to the data presenting high frequency, volatility, non-linearity, and seasonality [4]. Moreover, climatic variables, energy demand, power supply capacity, and the impact of renewable energy sources [5–7] make the forecasting process a challenging task. The incorporation of exogenous variables in the forecasting models could help the models to understand the data dynamics and allow for them to obtain more accurate results.

In Brazil, the electricity price can be approached as an optimization problem that assesses the level of the reservoirs of large hydroelectric plants. If the level of the reservoirs is low, water may be scarce in the future, depending on an analysis of stochastic forecasting [8]. When there is a tendency to reduce the level of the reservoir, the price of electricity is increased, both for industries in the short-term market and homes through the implementation of higher tariff flags. This procedure is used, because, in Brazil, there is a greater source of hydraulic generation, which requires adequate planning and control of the energy price to guarantee the supply of electricity reliably [9].

To develop accurate forecasting models, there is a trend that involves using combined or hybrid forecasting methods such as pre-processing (decomposition), optimization (single and multi-objective approaches), and artificial intelligence models [10–13]. Within this context, each methodology can add to the forecasting model its own expertise to deal with different signals characteristics. In this aspect, pre-processing approaches, especially decomposition methods, have the objective of filtering data noise and non-linearities, by decomposing the original signal into different kinds of frequencies (namely components) [14].

Each component can represent the trend, seasonality, high, and low frequencies. In this respect, in the forecasting field, a different class of models (heterogeneous ensemble learning of components) or the same model (homogeneous ensemble learning of components) regarding learning structure can be used to train and predict each decomposed component of the evaluated signal. Through this process, the diversity is enhanced and, when a final efficient forecasting model is obtained by aggregation (directly aggregation), an efficient model is obtained. Alongside this, evolutionary computation and swarm intelligence algorithms can be used to tune the hyperparameters of machine learning models [15] or the time series decomposition methods [16], aiming to make the model more accurate.

1.1. Related Works

When considering electricity price forecasting, previous studies have given attention to hybrid forecasting models. Firstly, Yang et al. [17] combined the decomposition methods variational mode decomposition (VMD) and improved complete ensemble empirical mode decomposition with adaptive noise. Subsequently, each component was trained and predicted by Elman partially recurrent neural network optimized by a multi-objective grey wolf optimizer. In the same way, Zhang et al. [18] used VMD with hyperparameters defined by self-adaptive particle swarm optimization. In order to forecast the modes in a one-step-ahead system, seasonal autoregressive integrated moving average and deep belief network were adopted to forecast regular and irregular modes, respectively.

Adjacent to the previous studies, Qiao and Yang [19] adopted wavelet transform coupled with stacked autoencoder model and long short-term memory (LSTM) to forecast residential, commercial, and industrial electricity prices. In its turn, Zhang et al. [20] proposed a cuckoo search-based feature selection integrated to singular spectrum analysis and support vector regression (SVR). The LSTM can have high accuracy for predicting chaotic time series, and can be superior to classic forecasting algorithms. In addition to the electricity price forecast, the application of LSTM has a good performance for photovoltaic forecasting, among other applications [21].

Yang et al. [22] adopted VMD algorithm, improved multi-objective sine cosine algorithm, and regularized extreme learning machine for multi-step electricity price forecasting. In this approach, the VMD approach is employed to extract the data features, such as high and lower frequency, but did not perform discussions about the feasibility of other features in the forecasting system.

Zhou et al. [23] coupled LSTM and ensemble empirical mode decomposition (EEMD) to forecasting electricity markets of Pennsylvania, New Jersey, and Maryland. Khalid et al. [24] proposed an optimized deep neural network framework to conduct electricity price forecasting based on the Jaya optimizer and LSTM approach. The main drawback of this proposed approach is related to the lack of the use of exogenous variables and use of decomposition approaches to cap the data variability. The use of artificial intelligence models, as well as non-linear models are attractive tools to make the forecasting system more robust, as observed in Ribeiro and Coelho [12] and Ribeiro et al. [13].

Many papers related to electricity forecasting are based on energy consumption. As exposed by Alipour et al. [25] renewable energy resources have an uncertainty of electric power generation, which can lead to problems in the electrical power systems. As presented by Kazemzadeh et al. [26] load forecast, it is one of the main base studies for the planning and operation of the expansion of the electric power system. In [27], an evaluation of the forecast is made for a project up to 2030, many models are covered in this work in order to show that there is a better performance depending on the algorithm.

Heydari et al. [28] proposed a composed model based on VMD, feature selection (selection of features related to hours of the day) in the Pennsylvania-New Jersey-Maryland and Spanish electricity markets. This paper lacks the discussion about the feasibility of the use of other features related to the electricity market. The authors argued that the obtained results outperform the results of some benchmarks, such as generalized regression and radial basis function neural networks.

In addition to electricity price and load forecasting, the use of artificial intelligence is powerful to assess the development of possible failures in the electrical system [29]. As presented by Stefenon et al. [30], the use of the wavelet transform reduces signal noise, improving the analysis of chaotic time signals. The results using the wavelet group method of data handling proved to be superior to well-consolidated algorithms as LSTM and adaptive neuro fuzzy inference system. Additionally, pre-processing techniques of time series are robust approaches that are widely used to de-noise the raw signal, and then enhance the forecasting accuracy, as approached in da Silva et al. [31].

Focused on the analysis of the electrical system about electricity generation in Brazil, many works evaluated the generation capacity concerning a hydroelectric source, as presented in Brito et al. [32] and Fredo et al. [33], or as a hydrothermal problem as discussed by Finardi et al. [34] and van Ackooij et al. [35]. Moreover, according to Silveira Gontijo and Azevedo Costa [36] in Brazil, there is a predominance of hydroelectric generation (73%), which makes the analysis of hydroelectric energy price forecasting an important field of study.

According to the above-related papers, there is no consensus about which decomposition method to employ in price forecasting analysis. Based on this, a different kind of decomposition, named complementary ensemble empirical mode decomposition (CEEMD), can be considered. In this context, the CEEMD, an extension of the EEMD method, has been applied in several fields of knowledge, such as crude oil price forecasting [37], short-term photovoltaic power generation forecasting [38], and detecting epileptic seizures in electroencephalogram [39].

Yeh et al. [40] proposed the CEEMD, in which the paired noises are perfectly anti-correlated and have an exact cancellation of the residue noise in the reconstruction of the signal. The CEEMD split the original signal in a set of components named intrinsic mode functions (IMFs) and one residue component, which are designed to represent a trend, seasonality, high and lower frequency. When considering the aforementioned, due to the use of the CEEMD approach, two questions emerge. The first one lies in which algorithm should be employed to train and forecast each component. The second refers to the selection of CEEMD's hyperparameters, namely the number of components, ensemble, and amplitude of standard deviation.

Alongside this, especially in Brazil, the evaluation of the energy price forecast is very important, when considering that the price of electricity has great influence due to the level of the large water reservoirs of hydroelectric plants. As the system is hydrothermal, when the reservoir level is low, thermoelectric plants are activated, depending on energy planning. The cost of generating thermoelectric plants is higher than that of hydroelectric plants, as there is a variable cost due to the

inputs that are needed to produce energy, so there can be a wide range of higher prices in dry periods. For this reason, evaluating the forecast of the future price, based on seasonal market and climate factors, generate greater reliability for the energy market.

1.2. Objective and Contribution

Therefore, this paper proposes a self-adaptive decomposed heterogeneous ensemble learning model. The methods CEEMD, coyote optimization algorithm (COA) [41] and machine learning are combined to develop a heterogeneous ensemble learning model, to forecasting commercial and industrial electricity prices in Brazil for multi-step-ahead (one, two, and three months-ahead) horizons. Exogenous variables, including energy generation (supply), energy prices lagged, and consumption (demand) are considered to be inputs by the forecasting model. Firstly, the COA optimizer is applied to define the CEEMD's hyperparameters and, subsequently, CEEMD decomposes the series of electricity energy prices (commercial and industrial). Thereafter, the components obtained in the previous step (IMFs and one residue component) are trained using extreme learning machines (ELM) [42], SVR [43], Gaussian process (GP) [44], and gradient boosting machines (GBM) [45]. These individual models are chosen due to the effects already observed for regression and time series forecasting tasks, as described in [46–48].

The hyperparameters of each model are obtained by grid-search during leave-one-out cross-validation time slice (LOOCV-TS). Finally, the prediction results of different components are directly integrated to generate the final electricity price. Afterwards, by the grid of models, the most adequate model is the one with the best generalization out-of-sample capacity in terms of coefficient of determination (R^2), symmetrical mean absolute percentage error (sMAPE), root mean squared error (RMSE), and overall weight average (OWA). The proposed framework is compared with COA-CEEMD homogeneous based methods, i.e., approaches which consider the same model to handle all components, as well as with homogeneous ensemble learning models that adopted maximal overlap discrete wavelet transform (MODWT) [49] for data pre-processing. Moreover, the autoregressive integrated moving average (ARIMA) [50], naïve [51], and theta models [52,53] are used as additional benchmarks.

The contributions of this paper are described, as follows:

- Firstly, this paper contributes to the field of time series pre-processing by coupling the CEEMD with metaheuristic approach named COA to tune its hyperparameters;
- Second, based on the literature review gap, exogenous variables related to supply and demand are used as inputs for each evaluated model, and their importance is assessed. The inputs associated to supply are the generation of hydraulic, nuclear, and thermal energy. Thus, the variables related to demand are the monthly consumption for each area (commercial and industrial). Through the use of these variables is intended to giving additional information for the models to learn the data behavior, so that they achieve high forecasting accuracy;
- Third, with the combination of the different non-linear models (ELM, SVR, GP, and GBM) to train and predict each component of the decomposed stage, the developed model can learn the data patterns and reflect the high-frequency of electricity price data; and,
- Also, this paper contributes for the literature of models used to forecasting electricity prices by investigating the performance of decomposed homogeneous and heterogeneous ensemble learning models.

The organization of the remainder of this paper is as follows: Section 2.1 presents the datasets that were adopted in this paper. Section 2.2 brings a brief description of the adopted methods in the forecasting framework. Section 3 details the procedures of the research methodology. Afterwards, Section 4 describes the results and discussions. Finally, Section 5 concludes the paper with final considerations, limitations of the study, and proposals of research directions.

2. Material & Methods

This Section presents the description of the data (Section 2.1) and methods applied in this paper (Section 2.2).

2.1. Material

The datasets analyzed in this paper refer to Brazil's commercial and industrial electricity prices (Brazilian currency—Real—R$) by megawatt-hour (MWh). Additionally, exogenous variables, such as energy generation (supply) and consumption (demand) (MWh), are considered. The datasets consist of 289 monthly observations from April 1996 to December 2019. These data were obtained from the website of the Institute of Applied Economics Research (IPEA) (*Instituto de Pesquisa Econômica Aplicada*, in Portuguese) available in http://www.ipeadata.gov.br/Default.aspx.

The datasets were split into training and testing sets in the proportion of 70% and 30%, respectively. The first 70% of them were used to train the adopted models and the remaining 30% of them were used to test the effectiveness of evaluated forecasting models. This range is commonly used in the literature, such as observed in Ribeiro and Coelho [12], and used in this paper. This proportion adopted to split the datasets into training and test sets allows for us to give the models more information to the adopted models learn the prices' dynamic, as well as to have a sufficient amount of the data to evaluate the effectiveness of the proposed model.

Figure 1 illustrates the electricity energy prices series, and the time series for the exogenous variables are shown in Figures 2 and 3. According to the Shapiro–Wilk normality test, the output variables for commercial and residential cases do not present a normal distribution (W = 0.9130 − 0.9215, *p*-value < 0.05). In Table 1, a summary of the statistical indicators of commercial and industrial electricity prices is shown. Moreover, previous prices are used as inputs signals by the forecasting system, and more details are described in the methodology section. However, to avoid repetition, the statistical indicators are not addressed in Table 1 for this variable.
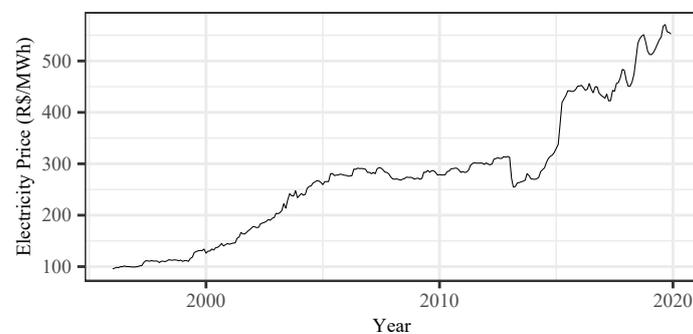

(**a**) Commercial
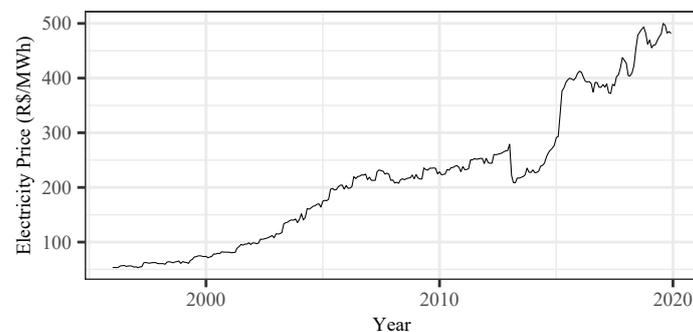
(**b**) Industrial

Figure 1. Brazilian electricity prices.

Table 1. Descriptive measures for Brazilian commercial, industrial electricity price, and exogenous variables.

Variable	Set	Dataset	Statistical Indicator				
			Minimun	Median	Mean	Maximum	Standard Deviation
Output (Price)	Whole	Commercial	95.63	276.62	274.83	570.80	123.04
	Training		95.63	254.01	215.50	314.14	75.74
	Test		254.58	441.55	414.16	570.80	97.69
	Whole	Industrial	53.13	217.03	215.70	500.05	125.31
	Training		53.13	160.81	152.52	265.07	73.22
	Test		208.54	391.88	364.10	500.05	92.58
Input (Demand)	Whole	Commercial	2647.00	5068.00	5312.33	8198.00	1629.53
	Training		2647.00	4214.00	4429.82	7037.00	1039.63
	Test		6454.00	7380.50	7385.21	8198.00	460.02
	Whole	Industrial	8753.00	13,602.00	12,914.72	15,886.00	2036.33
	Training		8753.00	12,017.00	12,329.83	15,853.00	2138.43
	Test		12,538.00	14,122.50	14,288.53	15,886.00	681.88
Input (Supply)	Whole	Hydraulic	20,593.00	31,434.50	31,300.98	43,604.00	4520.99
	Training		20,593.00	29,586.50	30,197.74	42,429.00	4441.69
	Test		27,940.00	33,484.00	33,892.31	43,604.00	3560.08
	Whole	Thermal	265.00	1833.00	3888.18	13181.00	3721.82
	Training		265.00	1462.50	1654.18	7158.00	1046.48
	Test		4569.00	9119.00	9135.48	13,181.00	2112.50
	Whole	Nuclear	0	1172.50	1003.05	1504.00	450.61
	Training		0	1068.00	880.25	1480.00	464.16
	Test		515.00	1395.00	1291.49	1504.00	236.90
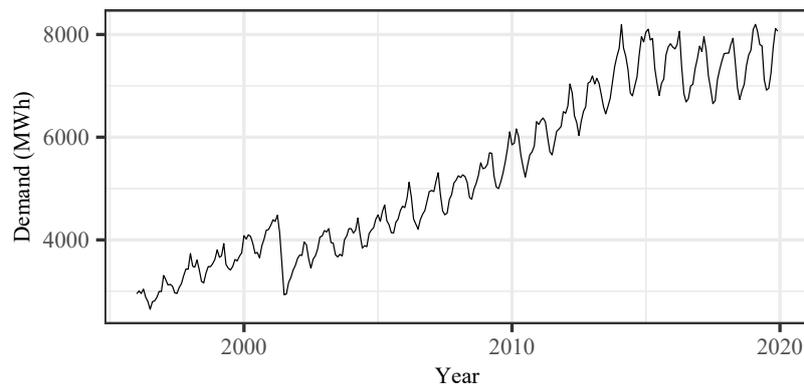
(**a**) Commercial
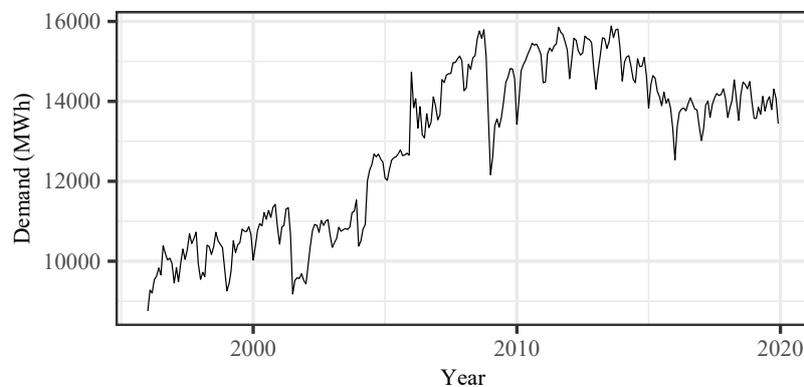
(**b**) Industrial

Figure 2. Brazilian electricity demand.
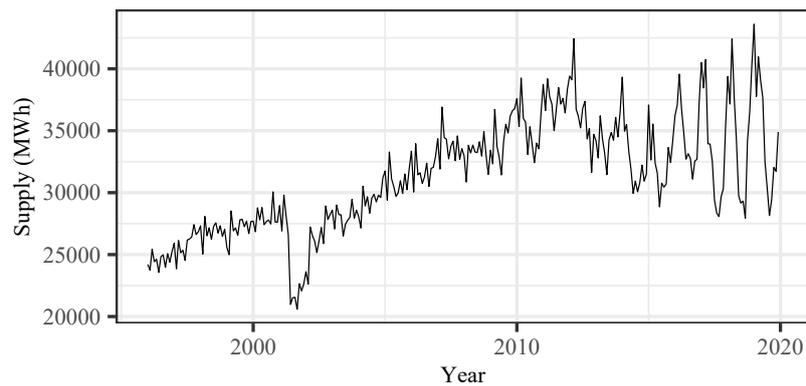
(**a**) Hydraulic
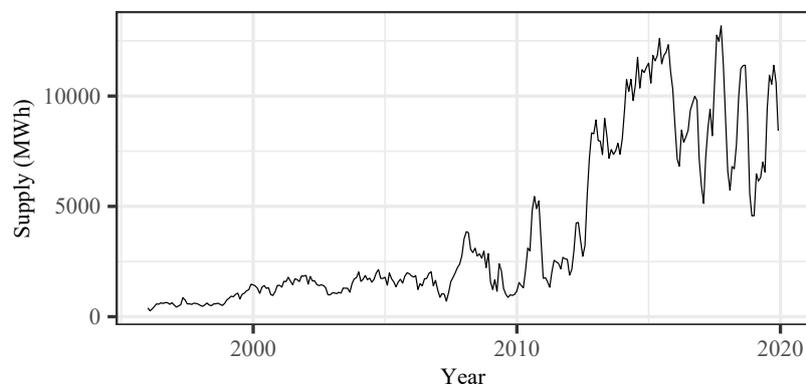
(**b**) Thermal
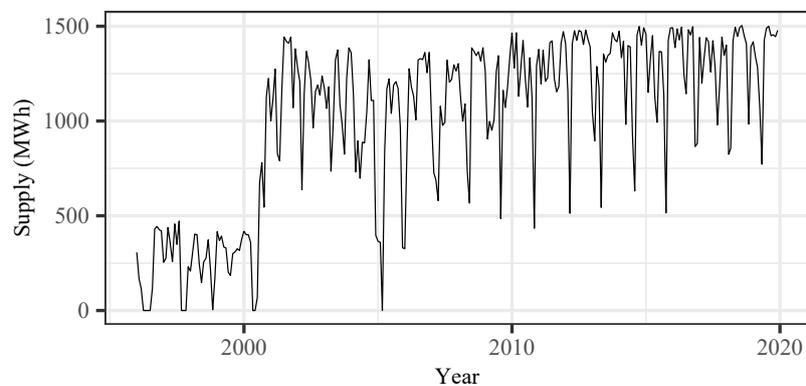
(**c**) Nuclear

Figure 3. Brazilian electricity supply.

2.2. Methods

This subsection describes the methods employed in this paper.

2.2.1. Coyote Optimization Algorithm

The COA optimizer is a swarm intelligence algorithm that considers the social relations of the *Canis latrans* species and its adaptation to the environment proposed by [41] devoted to solving optimization problems. Therefore, the COA mechanism has been designed based on the social conditions of the coyotes, which means the decision variables $\vec{x}$ of a global optimization problem [54].

The COA's performance was evaluated under 40 benchmark functions in the optimization field with different features as multimodality, nonlinearity, separability, and the number of optimized variables. In most of the benchmarks evaluated, set of 40 benchmark functions, the COA optimizer outperformed some classical metaheuristics, such as particle swarm optimization, artificial bee colony, symbiotic organisms search, grey wolf optimizer, bat-inspired algorithm, and firefly algorithm. Through these results, it is possible to state that COA could be applied in new applications and problems, such as that proposed in this paper.

This algorithm has recently been applied in several applications, especially to feature selection [54], tune heavy-duty gas turbine hyperparameters [55], optimal power flow for transmission power networks [56] define networks reconfiguration [57], and for optimal parameter estimation of a proton exchange membrane fuel cell [58]. Due to the promising potentials results, a search of the literature reveals that the COA has not yet been applied for the CEEMD's hyperparameters definition, then it is adopted. In the COA approach, there are only two control parameters, the number of packs (N_p) and the number of coyotes per pack (N_c). The population size is defined by multiplying (N_p) and (N_c), both natural numbers, and then the population is divided into packs with coyotes each.

2.2.2. Complementary Ensemble Empirical Mode Decomposition

The empirical mode decomposition (EMD) [59] and its improvements, such as EEMD and CEEMD [40], were proposed to deal with the non-linearity and non-stationarity of time series. The EMD separates the original signal into IMFs and one residue component. The main drawback of this decomposition is named mode mixed problem (MMP). The MMP is characterized by the fact that disparate scales could appear in one IMF. Next, to overcome this disadvantage, EEMD was proposed, and in the sequence CEEMD.

Despite the fact that EEMD has effectively resolved the MMP, the residue noise in the signal reconstruction has been raised, and the noise is independent and identically distributed [60]. To improve EEMD, [40] proposed the CEEMD, in which the paired noises are perfectly anti-correlated and have an exact cancellation of the residue noise in the reconstruction of the signal. Because of the effectiveness of CEEMD to de-noise time series, it has been applied for crude oil price forecasting [37], wind speed forecasting [61], and this paper employs this decomposition approach to pre-process the Brazilian electricity energy prices.

The CEEMD has three main hyperparameters, named as the number of trials or number of ensembles, the number of components, and noise amplitude. Especially, the noise amplitude is designed to be some percentage of the data standard deviation. In most of the cases, these hyperparameters are defined by trial and error procedure [37,60,62]. However, this paper proposes the use of the COA approach to minimize the inverse of the orthogonal index (OI) [16]. The OI is used to measure the orthogonality of the EMD numerically, and a value close to zero is desirable. A smaller OI indicates the best decomposition result [59].

The OI can be computed, as follows:

$$\mathrm{OI} = \sum_{t=0}^{T} \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{IMF}_i(t)\,\mathrm{IMF}_j(t)/x^2(t) \right), \tag{1}$$

in which T is the number of time-series observations, IMF_i and IMF_j are the i-th and j-th components, k is the number of components, and $x(t)$ is the original signal at time $t = 0, \dots, T$.

2.2.3. Extreme Learning Machine

The ELM is a learning algorithm proposed by [42] designed for single-hidden layer feedforward neural networks. In this approach, hidden nodes are randomly chosen and outputs are obtained analytically. Good generalization and fast learning speed are the main advantages of ELM [48]. The input weights and hidden biases are specified arbitrarily and then are fixed. The output weights

are obtained by solving the multiplication of the Moore–Penrose Generalized inverse matrix of the output variable matrix [63].

2.2.4. Gradient Boosting Machine

The GBM is an ensemble learning approach that employs a sequential learning process to build an efficient classification or regression model [45]. A regression tree is initially fitted to the data and, on this basis, predictions and the initial residue are computed. A new model is fitted to the previous residue, a new prediction, to which the initial forecast is added, and then a new residue is obtained. This process is iteratively repeated until a convergence criterion is obtained. In each iteration, a new model is fitted to the data, aiming to compensate for the weaknesses of the previous model [12].

2.2.5. Gaussian Process

A GP is a stochastic process, in which every set of the random variable is multivariate normally distributed. In this respect, a GP is entirely specified by its statistical orders mean and covariance or kernel function. Through kernel function, it is possible to maps the similarity between observations of the training set with the purpose of predicting new observations [44].

2.2.6. Support Vector Regression

The SVR consists in determining support vectors close to a hyperplane that maximizes the margin between two-point classes obtained from the difference between the target value and a threshold. To deal with non-linear problems, SVR takes into account kernel functions, which calculates the similarity between two observations. In this paper, the linear kernel is adopted. The main advantages of the use of the SVR lie in its capacity to capture the predictor non-linearity and then use it to improve the forecasting cases. In the same direction, it is advantageous to employ this perspective in this adopted case study, once the samples are small [64].

2.3. Performance Indicators

To check the forecasting models' performance, the sMAPE (2), R^2 (4), and RMSE (3) criteria are used. Additionally, a criterion which combines the sMAPE and RMSE, named OWA (5) is introduced to evaluating the accuracy of the proposed model against benchmark compared models. According to Makridakis et al. [65], the use of OWA would help to achieve a higher level of objectivity. These measures are described in the following.

$$\text{sMAPE} = \frac{2}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}, \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}, \tag{4}$$

$$\text{OWA} = \frac{1}{2} \left[\frac{\text{sMAPE}_{\text{p}}}{\text{sMAPE}_{\text{c}}} + \frac{\text{RMSE}_{\text{p}}}{\text{RMSE}_{\text{c}}} \right]. \tag{5}$$

where n is the number of observations, y_i and $\hat{y}_i$ are the i-th observed and predicted values, respectively. additionally, the Criteria$_{\text{c}}$ and Criteria$_{\text{p}}$ represent the performance measure of compared and proposed model, respectively.

By considering the criteria sMAPE, and RMSE, lower values are desired, while, for R^2, a value closest to one indicates better performance. Additionally, the Diebold–Mariano (DM) test [66] is applied in this paper to compare the forecasting errors of proposed versus compared forecasting models.

3. The Proposed Self-Adaptive Decomposed Heterogeneous Ensemble Learning Model

This section presents the steps adopted to develop the self-adaptive decomposed heterogeneous ensemble learning model.

Step 1: The COA is coupled with CEEMD to tune the CEEMD's hyperparameters. For COA optimizer, the number of coyotes and packs are defined as 5 and 10, respectively. These values are selected by the trial and error, once that there is no guideline for the definition of COA's hyperparameters [41]. Moreover, if the increase in the number of packs, and/or coyotes is considered, the optimization time will also increase due to the greater number of evaluations to be carried out. However, for this problem, it was observed that the accuracy does not improve significantly. In this way, the initial values adopted for these parameters are fixed, for both problems.

Table 2 shows the CEEMD's hyperparameters defined by COA, where 50 generations and population of size equals to 50 is adopted. Finally, the original electricity price is decomposed.

Table 2. Search boundaries of the CEEMD hyperparameters.

Hyperparameter	Boundaries		Selected Hyperparameters	
	Lower Bound	Upper Bound	Commercial	Industrial
Number of ensembles	50	100	51	85
Number of Components	2	5	4	4
Noise amplitude	0.2	0.5	0.4049	0.3134

Step 2: Each component obtained in step 1 (three IMFs and one residue) is trained using ELM, GBM, GP, and SVR. In the training stage, LOOCV-TS IS adopted. The inputs are defined by auto-correlation and partial auto-correlation analysis. The data are centered by its mean value and divided by its standard deviation. The training structure is stated, as follows:

$$y_{(t+1,k)} = f\left\{y_{(t,k)}, \dots, y_{(t-n_y,k)}, \mathbf{X}_{(t+h-n_x)}\right\} + \epsilon \quad \epsilon \sim N(\mathbf{0}, \sigma^2), \tag{6}$$

and forecast electricity energy prices one-month-ahead (7), two-months-ahead (8), and three-months-ahead (9) according to:

$$\hat{y}_{(t+h,k)} = f\left\{y_{(t+h-1,k)}, y_{(t+h-2,k)}, y_{(t+h-3,k)}, \mathbf{X}_{(t+h-1)}\right\} \tag{7}$$

$$\hat{y}_{(t+h,k)} = f\left\{\hat{y}_{(t+h-1,k)}, y_{(t+h-2,k)}, y_{(t+h-3,k)}, \mathbf{X}_{(t+h-2)}\right\} \tag{8}$$

$$\hat{y}_{(t+h,k)} = f\left\{\hat{y}_{(t+h-1,k)}, \hat{y}_{(t+h-2,k)}, y_{(t+h-3,k)}, \mathbf{X}_{(t+h-3)}\right\} \tag{9}$$

in which f is a function that is related to the adopted model in the training stage, $\hat{y}_{(t+h,k)}$ is the forecast value for k-th component obtained in the decomposition stage ($k = 1, \dots, 4$) on time t and forecast horizon h ($h = 1, 2, 3$), $y_{(t+h-n_y,k)}$ are the previously prices lagged in $n_y = 1, \dots, 3$ and ϵ is the random error which follows a normal distribution N with zero mean and variance σ^2. Also, $\mathbf{X}_{(t+h-n_x)}$ is the exogenous inputs vector at the maximum lag of inputs ($n_x = 1$ if $h = 1$, and $n_x = 3$ if $h = 3$).

The behavior observed in the residue component of both time series is due to the growing trend in prices. Additionally, for the GP approach, there are no hyperparameters for tuning and the linear kernel function is used. Moreover, for GBM the shrinkage and minimum number of terminal node size are held constant equal 0.1 and 10, respectively. Finally, for SVR, the Radial Basis kernel

function is adopted. The kernels of GP and SVR were defined by grid-search by RMSE minimization during LOOCV-TS.

Step 3: The forecasts of different models used for each component are directly integrated (simple sum) to generate final electricity price values. Afterwards, by the grid of models, the most adequate model is the one with the best generalization out-of-sample capacity in terms of RMSE and sMAPE. Table 3 describes the models that are used for each component.

Table 3. Models adopted by each component in each dataset.

Dataset	Component	Forecasting Horizon		
		One-Month-Ahead	Two-Months-Ahead	Three-Months-Ahead
Commercial	IMF_1	GBM	GBM	ELM
	IMF_2	ELM	ELM	GBM
	IMF_3	SVR	SVR	GP
	Residue	SVR	SVR	GP
Industrial	IMF_1	GBM	GBM	GP
	IMF_2	GBM	GP	GP
	IMF_3	GP	GP	GBM
	Residue	GP	GP	SVR

The decomposed commercial and industrial electricity prices are illustrated in Figures 4 and 5, respectively.
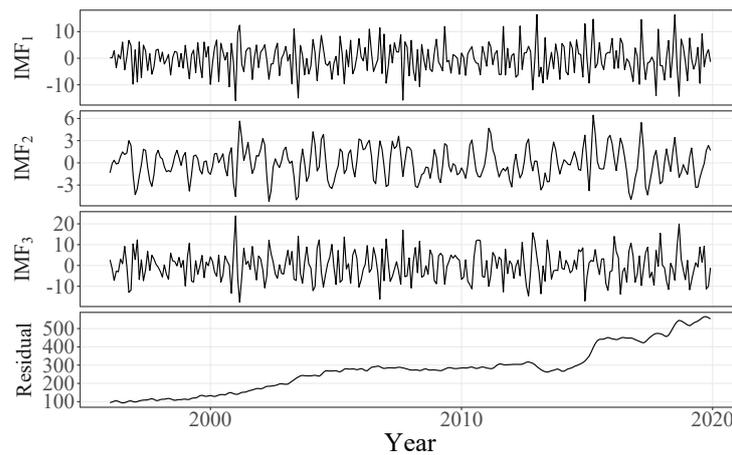

**Figure 4.** Decomposed series for commercial prices.
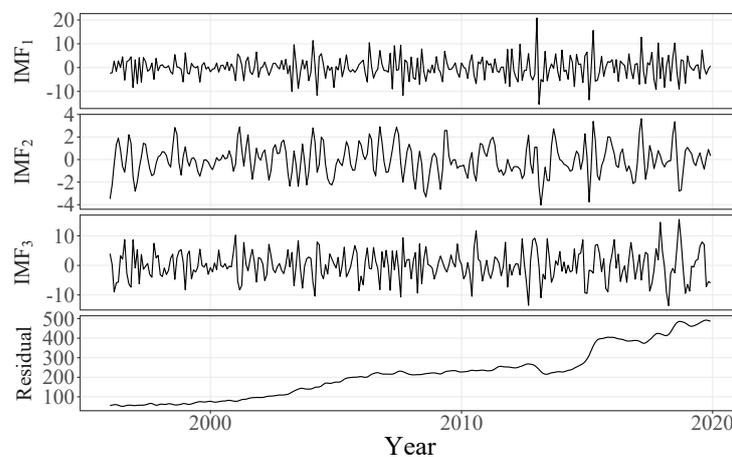
Figure 5. Decomposed series for industrial prices.

Step 4: Obtaining forecasts out-of-sample (test set), and the performance indicators defined in Section 2.3 are computed and two kinds of comparisons are conducted. The first is the comparison of self-adaptive decomposed heterogeneous and homogeneous ensemble learning models. Second, a comparison of the proposed model and models without decomposition are developed.

Table 4 presents the models' hyperparameters obtained by grid-search.

Table 4. Hyperparameters selected by grid-search for each adopted approach.

Dataset	Component	Forecasting Horizon	ELM			SVR	GBM	
			# Neurons	Activation Function	Weights Initialization	Cost	Boosting Interactions	Maximum Tree Deph
Commercial	IMF$_1$	One-month-ahead	12	Sigmoide	Uniform Positive	0.25	50	1
		Two-months-ahead	8	Tribas	Uniform Negative	1	50	1
		Three-months-ahead	8	Satlins	Uniform Negative	1	50	1
	IMF$_2$	One-month-ahead	8	Relu	Uniform Positive	0.25	150	3
		Two-months-ahead	5	Hardlin	Uniform Negative	0.5	50	3
		Three-months-ahead	5	Sigmoide	Uniform Negative	0.25	50	3
	IMF$_3$	One-month-ahead	3	Radial Basis	Uniform Negative	0.25	50	1
		Two-months-ahead	8	Hardlin	Normal Gaussian	0.25	150	2
		Three-months-ahead	12	Sine	Uniform Negative	0.25	50	1
	Residue	One-month-ahead	8	Sigmoide	Uniform Negative	0.5	150	3
		Two-months-ahead	12	Sigmoide	Uniform Negative	0.5	150	3
		Three-months-ahead	12	Sigmoide	Uniform Negative	0.5	150	3
	Non-Decomposed	One-month-ahead	12	Sigmoide	Uniform Negative	1	150	3
		Two-months-ahead	12	Sigmoide	Uniform Negative	1	150	3
		Three-months-ahead	12	Sigmoide	Uniform Negative	1	150	3
Industrial	IMF$_1$	One-month-ahead	12	Purelin	Uniform Positive	1	50	2
		Two-months-ahead	8	Relu	Uniform Positive	1	100	3
		Three-months-ahead	15	Satlins	Uniform Negative	0.25	150	2
	IMF$_2$	One-month-ahead	12	Purelin	Uniform Positive	0.25	50	2
		Two-months-ahead	8	Relu	Uniform Positive	0.25	50	2
		Three-months-ahead	8	Relu	Uniform Positive	0.25	50	2
	IMF$_3$	One-month-ahead	8	Sigmoide	Uniform Negative	1	50	1
		Two-months-ahead	8	Sigmoide	Uniform Positive	0.25	50	1
		Three-months-ahead	3	Tansig	Uniform Negative	0.25	50	1
	Residue	One-month-ahead	5	Sigmoide	Uniform Positive	1	150	3
		Two-months-ahead	5	Sigmoide	Uniform Positive	1	150	3
		Three-months-ahead	5	Sigmoide	Uniform Positive	1	150	3
	Non-Decomposed	One-month-ahead	5	Sigmoide	Uniform Positive	0.25	150	3
		Two-months-ahead	5	Sigmoide	Uniform Positive	1	100	3
		Three-months-ahead	1	Sigmoide	Uniform Positive	1	150	2

Figure 6 summarize the main steps used in data analysis.
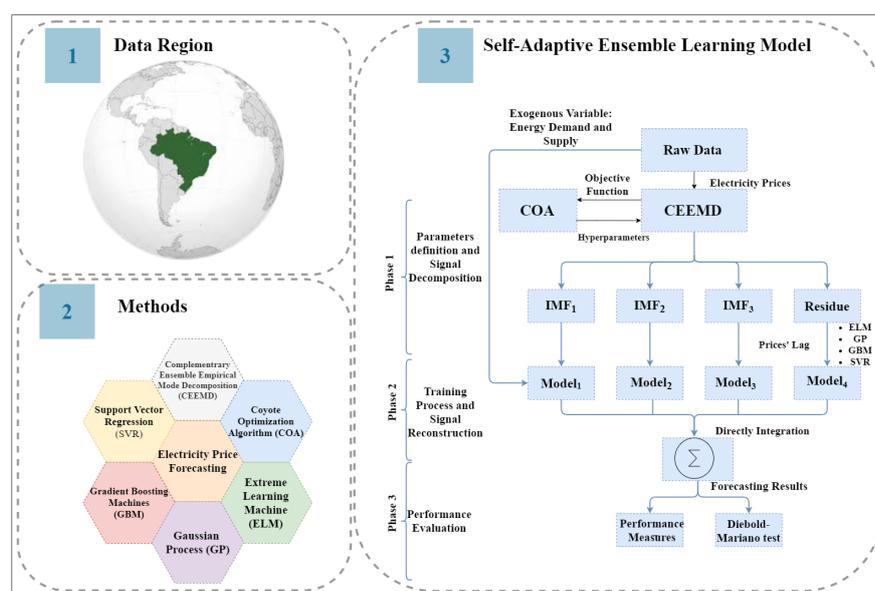
Figure 6. Framework of proposed approach.

The R software [67] is adopted to perform the modeling. The packages `caret` [68] and `forecTheta` [69]. The ARIMA modeling is performed through the use of `forecast` package [70,71] with use of `auto.arima` function. To define the ARIMA order, grid-search is adopted, and the most suitable order is that reach lower Akaike and Bayesian Akaike criteria information.

4. Results

This section describes the results of the developed experiments in three ways in forecasts out-of-sample (test set). First, Section 4.1 is designed to compare the results of the proposed model and self-adaptive decomposed homogeneous ensemble learning models. In the sequence, Section 4.2 is used to compare the performance of developed approaches and non-decomposed models. To finish, Section 4.3 presents the DM test to statistically evaluate the errors of the proposed approach versus other models. Additionally, Figure 7 presents the variables importance, Figures 8 and 9 illustrate the relation between the observed and predicted values. Additionally, Figure 10 shows the magnitude of the sum of standardized squared errors. In Tables 5 and 6, the best results are presented in bold.
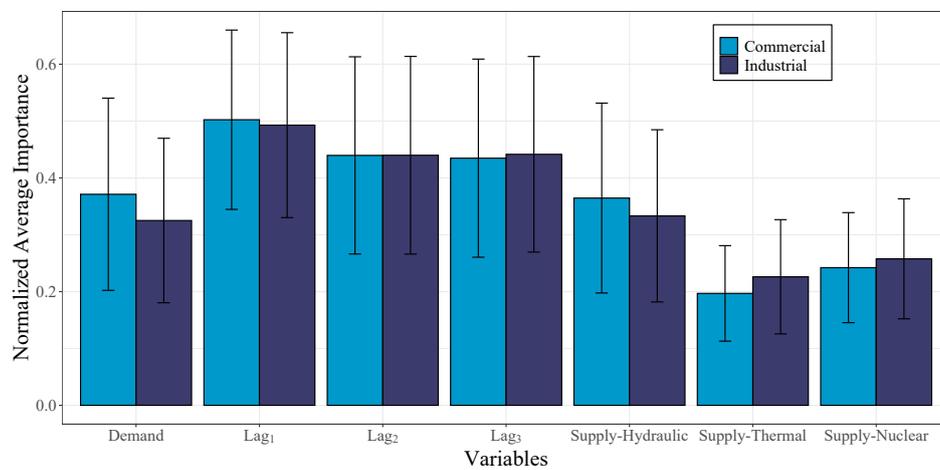

**Figure 7.** Variables Importance for the proposed model in each dataset.
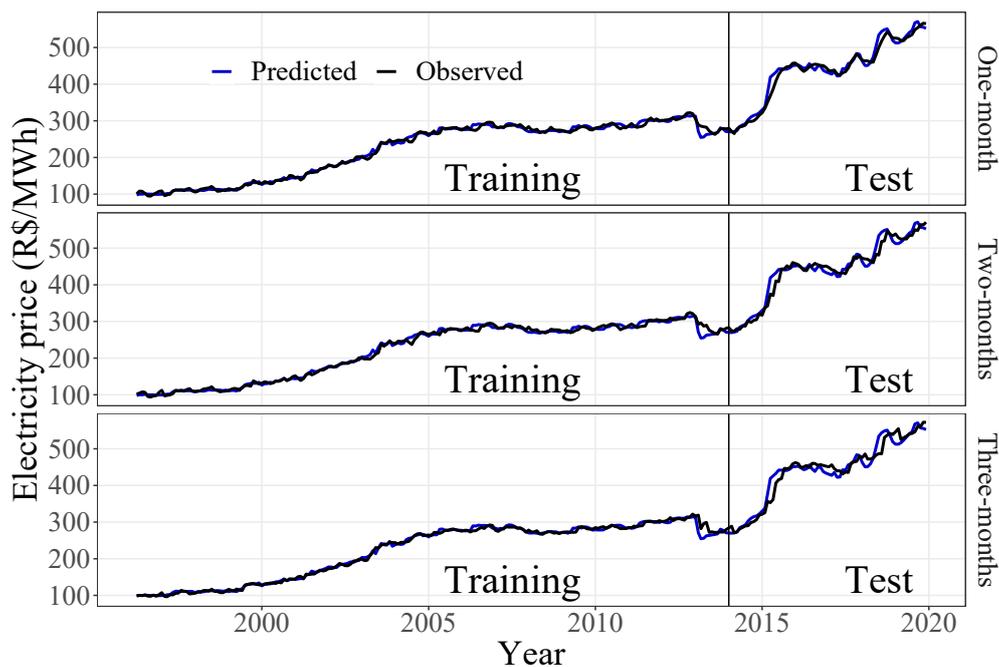
Figure 8. Predicted and observed Brazilian commercial electricity prices.
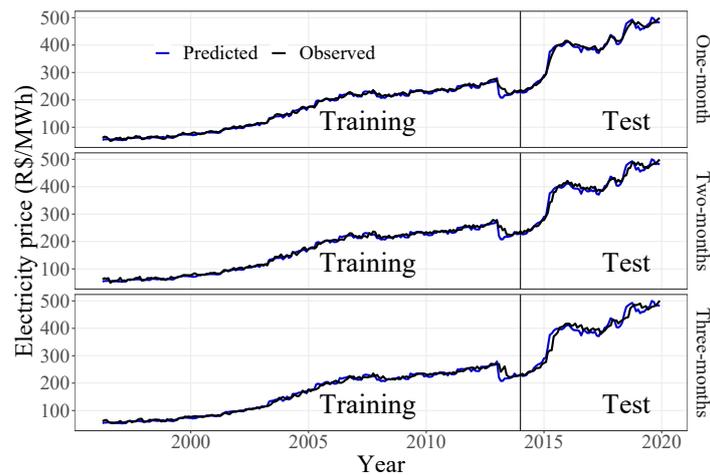
Figure 9. Predicted and observed Brazilian industrial electricity prices.

(**a**) Commercial　　　　　　(**b**) Industrial
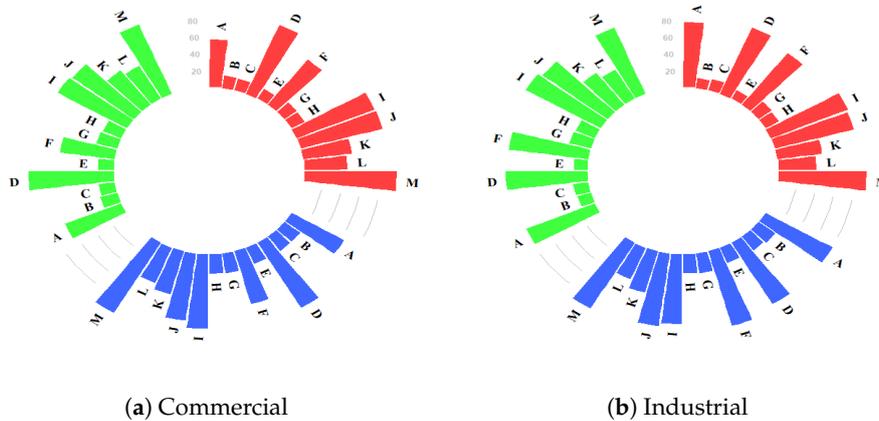
Figure 10. Radar plot for standardized sum of squared errors.

Table 5. Performance Measures of proposed and homogeneous ensemble learning models.

| Dataset | Model | Forecasting Horizon | | | | | | | | |
| | | One-month-Ahead | | | Two-Months-Ahead | | | Three-Months-Ahead | | |
		RMSE	R^2	sMAPE	RMSE	R^2	sMAPE	RMSE	R^2	sMAPE
	COA-CEEMD–Proposed	**13.5556**	**0.9812**	**0.0253**	**16.8360**	**0.9701**	**0.0306**	**20.7988**	**0.9544**	**0.0380**
	COA-CEEMD–GP	14.2734	0.9798	0.0260	17.2289	0.9692	0.0313	21.0945	0.9531	0.0382
	COA-CEEMD–SVR	16.1779	0.9774	0.0315	18.7674	0.9652	0.0352	21.5886	0.9516	0.0387
	COA-CEEMD–ELM	123.9228	0.6636	0.3032	122.1742	0.7888	0.2660	126.9495	0.7244	0.2871
Commercial	COA-CEEMD–GBM	143.4820	0.5415	0.2971	143.8438	0.4960	0.3001	143.9894	0.5081	0.3021
	MODWT–GP	22.9283	0.965	0.0412	31.4405	0.94	0.0571	35.6375	0.924	0.066
	MODWT–SVR	23.6391	0.969	0.0444	29.5238	0.948	0.0571	34.2929	0.92	0.0687
	MODWT–ELM	106.667	0.852	0.2512	120.765	0.636	0.2832	151.884	0.638	0.4098
	MODWT–GBM	145.605	0.398	0.3055	146.527	0.355	0.3078	145.6	0.367	0.3068
	COA-CEEMD–Proposed	**11.5992**	**0.9849**	**0.0256**	**14.8531**	**0.9750**	**0.0327**	**19.7095**	**0.9544**	**0.0418**
	COA-CEEMD–SVR	12.0007	0.9844	0.0252	15.1032	0.9741	0.0341	20.4148	0.9510	0.0432
	COA-CEEMD–GP	16.4187	0.9785	0.0368	16.7976	0.9694	0.0351	28.8510	0.9460	0.0606
	COA-CEEMD–ELM	127.7698	0.6352	0.3150	128.7123	0.6397	0.3187	130.8896	0.4547	0.3248
Industrial	COA-CEEMD–GBM	143.0262	0.4078	0.3467	141.0850	0.4990	0.3420	140.7303	0.4346	0.3443
	MODWT–SVR	31.0218	0.964	0.0697	52.2598	0.887	0.1164	52.3042	0.859	0.1159
	MODWT–GP	32.647	0.958	0.0732	40.9644	0.915	0.0898	42.0218	0.892	0.0973
	MODWT–ELM	93.2804	0.634	0.2387	110.765	0.697	0.2752	123.604	0.496	0.3069
	MODWT–GBM	141.895	0.357	0.3503	142.396	0.325	0.3534	143.233	0.227	0.3583

Table 6. Performance measures of proposed and compared models used to forecasting Brazilian electricity price with multi-step-ahead

Dataset	Model	Forecasting Horizon								
		One-Month-Ahead			Two-Months-Ahead			Three-Months-Ahead		
		RMSE	R^2	sMAPE	RMSE	R^2	sMAPE	RMSE	R^2	sMAPE
	COA–CEEMD–Proposed	**13.5556**	**0.9812**	**0.0253**	**16.8360**	**0.9701**	**0.0306**	**20.7988**	**0.9544**	**0.0380**
	GP	16.5411	0.9725	0.0294	21.8210	0.9572	0.0388	24.6019	0.9427	0.0438
Commercial	SVR	17.6207	0.9710	0.0324	25.3748	0.9537	0.0489	23.7418	0.9430	0.0397
	ELM	121.6424	0.7708	0.2559	126.6971	0.7882	0.2879	133.8389	0.7073	0.3095
	GBM	143.9112	0.5083	0.2978	145.1188	0.4954	0.3034	143.3755	0.5001	0.2986
	COA–CEEMD–Proposed	**11.5992**	**0.9849**	**0.0256**	**14.8531**	**0.9750**	**0.0327**	**19.7095**	**0.9544**	**0.0418**
	SVR	16.8467	0.9692	0.0335	20.4888	0.9507	0.0401	23.6189	0.9374	0.0489
Industrial	GP	21.0049	0.9642	0.0445	24.6607	0.9417	0.0541	23.2372	0.9404	0.0466
	ELM	127.2140	0.6150	0.3146	128.3130	0.6206	0.3166	145.2424	0.0236	0.3718
	GBM	148.5723	0.3560	0.3680	145.9830	0.4461	0.3610	142.8180	0.5746	0.3506

4.1. Comparison of Proposed and Self-Adaptive Decomposed Homogeneous Ensemble Learning Model

Table 5 illustrates the performance of developed and self-adaptive decomposed homogeneous ensemble learning models named COA-CEEMD-GP, COA-CEEMD-ELM, COA-CEEMD-SVR, and COA-CEEMD-GBM, as well compared MODWT-based homogeneous models.

By investigating the improvement on the errors of the proposed model regarding compared COA-CEEMD homogeneous ensemble learning models, it is possible to infer that the OWA criterion is ranged between 2.51–45.28%, 2.23–89.05%, and 0.87–86.48%, for commercial electricity price on one, two and three-months-ahead forecasting, respectively. In the comparison of COA-CEEMD heterogeneous ensemble learning model with MODWT ensemble learning models, the OWA ranges between 20.44–45.35%, 44.69–89.28%, and 42.02–88.51%, for commercial electricity price on one, two, and three-months-ahead forecasting, respectively.

The reason for the high forecasting error of COA-CEEMD-GBM and COA-CEEMD-ELM is attributed to forecasting errors and its high variability for each component. In general, the proposed model employed ELM or GBM for the first two components in the context of adopted datasets, as mentioned in Table 3. For these components, the GBM and ELM models achieved lower forecasting errors.

When considering the COA-CEEMD-GBM and COA-CEEMD-ELM models for the commercial dataset, the GBM and ELM models have achieved a high forecasting error values for the components 3 and 4. A similar analysis can be conducted in the context of industrial dataset. Therefore, this results lead the models to achieve a high forecasting error in the general model.

In the industrial context, the same pattern is observed, once the regarding the MODWT based models is ranged between 62.95–92.26%, 63.65–90.15%, and 55.06–87.29% in horizons of one up to three-months-ahead, respectively.

Therefore, it should be noted that the use of different models to compose an ensemble of components improves the final accuracy of the forecasting model. This is verified for the case of commercial as well as industrial electricity prices. The observed superiority of the proposed approach regarding these compared models should be attributed to the fact that diversity of heterogeneous ensembles is higher than homogeneous ensembles, which plays an important role in ensemble learning model [72].

4.2. Comparison of Proposed and Non-Decomposed Models

Table 6 shows the performance of the developed and non-decomposed models, named GP, ELM, SVR, and GBM.

Concerning the enhancement of hybrid model regarding non-decomposed models, the OWA in one-month-ahead forecasting ranged between 9.02–45.02% and 27.36–92.62% in commercial and industrial electricity prices, respectively. Similar behavior is observed for the other two time windows. In fact, the OWA ranges between 42.02–88.51% and 22.91–90.28% for two-months-ahead horizon,

as well as 8.25–86.38% and 12.71–87.59%, for three-months-ahead horizon. The performance of the proposed ensemble learning model is excellent, once that the forecasting errors are lower than 5%.

In respect of the non-decomposed approaches, the GBM and ELM models have difficulty in training and predicting the electricity prices due to high level of uncertain of the prices, which justify the lower performance. In this way, results lead to high forecasting errors and lower stability as point out in Tables 5 and 6, as well as in Figure 10.

In respect of the results thatare presented in Table 6, the outcomes of this paper reinforce findings presented by Hao et al. [73], which point out the benefits of using decomposition techniques as a way of pre-processing time series. In particular, the use of the COA-CEEMD approach is important for the development of an effective model for forecasting electricity prices. Second, to Yeh et al. [40], the use of signal decomposition as the pre-processing step is useful in the analysis of time series field because through the use of this technique it is possible to deal with non-stationarity and non-linearity behaviors of the data. Additionally, the results that are described in this section corroborate the findings of Agrawal et al. [74], once the ensemble learning models achieved better accuracy than its members.

In comparison to the proposed model with the ARIMA model, the naïve, and theta models, there are the following results. When considering the average (standard deviation) of RMSE for the proposed model over the three forecasting horizons, the average accuracy equals 17.05 (2.96) and 15.38 (3.33) in the commercial and industrial datasets, respectively. When considering the ARIMA model, the average accuracy equals to 22.24 (5.02) and 24.06 (4.28), in the commercial and industrial datasets, respectively. In respect to the naïve method, the average accuracy equals 23.84 (5.14) and 24.30 (4.54), in the commercial and industrial datasets, respectively. Finally, for the theta model, the average accuracy equals to 23.50 (5.06) and 24.25 (4.35), in the commercial and industrial datasets, respectively. In the context of described accuracy, the proposed model outperforms the results of these three benchmarks approaches. The naïve model forecasting the next h steps-ahead equal to the previous h time steps, which do not address satisfactory information to the decision making process. The ARIMA, naïve, and theta models are less complex than the proposed model, and they also achieved competitive results. However, these approaches are not adequate for forecasting the time series adopted in this paper (single input and multiple-output). Moreover, they do not allow for incorporating the information of exogenous variables such as supply and demand, as well as to access the importance of these features.

4.3. Statistical Tests to Compare Proposed and Benchmark Models

In order to demonstrate the statistical comparisons between errors of the proposed and compared models described in Sections 4.1 and 4.2, in Table 7 can be seen the statistic of DM test, as well as when the comparisons are statistically significant.

Table 7. Statistic of DM test for statistical comparison of proposed approach versus other models

| Model | Forecasting Horizon | | | | | |
| | One-Month-Ahead | | TwO-Months-Ahead | | ThRee-Months-Ahead | |
	Commercial	Industrial	Commercial	Industrial	Commercial	Industrial
COA-CEEMD–ELM	−10.63 ***	−9.88 ***	−5.68 ***	−5.70 ***	−4.61 ***	−4.41 ***
COA-CEEMD–SVR	−3.53 ***	−0.64	−1.65 *	−1.09 *	−1.52 *	−2.72 **
COA-CEEMD–GP	−1.75 **	−3.69 ***	−0.80	−1.25 *	−0.65	−1.32 *
COA-CEEMD–GBM	−9.37 ***	−9.95 ***	−5.45 ***	−5.89 ***	−4.18 ***	−4.45 ***
MODWT–ELM	−11.10 ***	−8.33 ***	−6.07 ***	−5.84 ***	−6.66 ***	−4.37 ***
MODWT–SVR	−−5.34 ***	−7.03 ***	−3.51***	−5.94 ***	−2.73 ***	−3.84 ***
MODWT–GP	−5.13 ***	−6.99 ***	−3.84 ***	−4.94 ***	−3.18 ***	−4.52 ***
MODWT–GBM	−9.49 ***	−10.15 ***	−5.45 ***	−5.89 ***	−4.14 ***	−4.47 ***
ELM	−9.06 ***	−9.85 ***	−5.17 ***	−6.32 ***	−4.94 ***	−4.59 ***
SVR	−3.20 ***	−4.29 ***	−5.34 ***	−7.86 ***	−2.10 **	−1.67 *
GP	−2.55 **	−2.88 **	−6.10 ***	−7.06 ***	−1.33 *	−2.20 ***
GBM	−9.42 ***	−10.21 ***	−5.51 ***	−6.48 ***	−4.19 ***	−4.51 ***

Note: *** 1% significance level; ** 5% significance level; * 10% significance level.

Through the DM test, it can be stated that, in 95.83% of the cases, the proposed approach reached statistically lower errors than the other models. The GP model reaches similar errors regarding the proposed forecasting framework, but greater than proposed modeling.

Figure 7 illustrates the importance of each variable to the forecasting proposed models.

It is possible to observe that the past electricity prices are the most important. In the next, the demand variables (commercial and industrial consumption) and supply hydraulic energies have similar importance. Finally, thermal and nuclear supply are less important, but they should be considered by the forecasting model.

Figures 8 and 9 expose that the self-adaptive decomposed ensemble learning model learn the data behavior, being able to obtain forecast prices that are similar to the observed values. For commercial and industrial datasets, the good performance (regarding RMSE and sMAPE) in the training set is maintained in the test set.

Finally, Figure 10a,b present the standard error of each model in the out-of-sample forecasting for the adopted forecasting horizons, i.e., red (one-month-ahead), blue (two-months-ahead), and green (three-months-ahead). Moreover, each letter represents one model, i.e., COA-CEEMD-ELM (A), COA-CEEMD-SVR (B), COA-CEEMD-GP (C), COA-CEEMD-GBM (D), proposed (E), ELM (F), SVR (G), GP (H), GBM (I), MODWT-ELM (J), MODWT-SVR (K), MODWT-GP (L), and MODWT-GBM (M). The best models are represented by smaller bars. The models with better accuracy presented in Tables 5 and 6 reached better stability (errors with lower standard deviation), especially the proposed COA-CEEMD heterogeneous ensemble learning model. Therefore, the results that are exposed in previous sections are confirmed.

Based on Figure 11 it is possible to evaluate the residue (partial auto-correlation function—PACF) of final models adopted in each case study, from training set results. In this respect, once most of the lags are inside the boundaries, we can see that the models are well fitted.


(**a**) Commercial   (**b**) Industrial
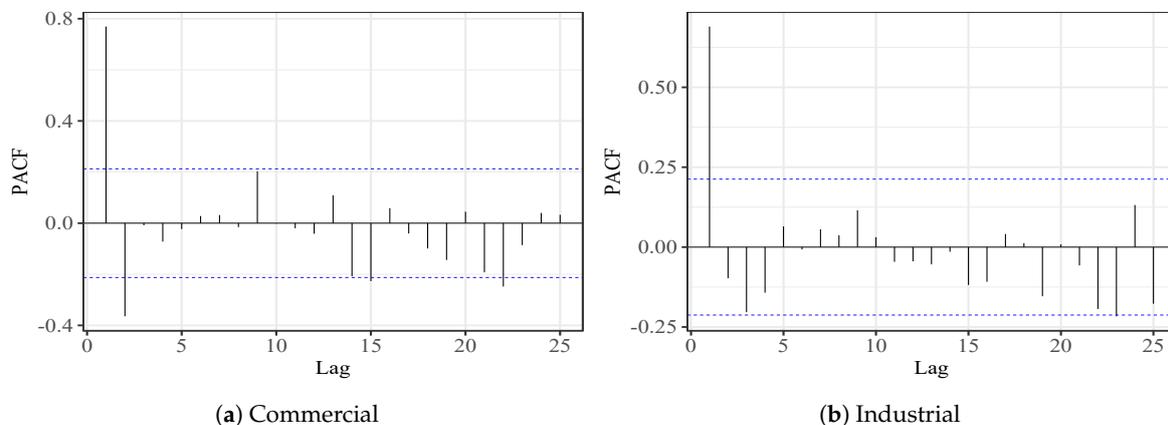
Figure 11. Partial autocorrelation for residue of proposed model in each dataset.

5. Conclusions

In this paper, a self-adaptive decomposed heterogeneous ensemble learning model was proposed in order to forecast multi-step-ahead (one, two, and three-months-ahead) Brazilian commercial and industrial electric energy prices. Exogenous variables, such as demand (commercial and industrial consumption) and supply (hydraulic, thermal, and nuclear) (MWh), were adopted. In the first stage, the COA optimizer was adopted to define the hyperparameters of pre-processing CEEMD. In the sequence, the four obtained components (three IMFs and one residue component) by COA-CEEMD were trained and predict the time series by different forecasting models (ELM, GBM, GP, and SVR). The grid-search approach was conducted in order to choose the most suitable model for each component. The final forecasts were obtained through a heterogeneous ensemble learning

of components directly integrated. Finally, the average importance of each variable used as model inputs was computed for the proposed forecasting model.

Our findings suggest that: (i) the COA-CEEMD ensemble learning models achieve better forecasting accuracy than single forecasting models; (ii) the pre-processing of the energy prices through COA-CEEMD is better than MODWT; (iii) the use of different models for components allow for improving the final accuracy concerning the use of a homogeneous ensemble model of components; (iv) the proposed approach reaches better accuracy than the compared models, and the good performance is constant when the forecast horizon is expanded; and, (v) the variables importance ranking, from the smallest to the larger is, energy prices lagged, demand, and supply. Moreover, competitive results are achieved by ARIMA, naïve, and theta models regarding the proposed framework. However, these models do not allow us to evaluate the importance of exogenous inputs, which plays a key role in the decision-making process.

The evaluation of the electricity price forecast can be used to assess feasibility for future expansions of the electric system, through the incentive of investment in the electric sector using incentive policies. Defining how much energy may cost in the future can facilitate the justification for investing in new electricity generation ventures. In Brazil, there is a potential for the use of wind farms, due to the geographic characteristics of the country. The evaluation of the investment potential in this segment can be analyzed based on the forecast of future water inflow, when considering that the cost of energy is calculated based on the level of the water reservoirs.

For future works, it is desirable (i) to decompose the raw data using an ensemble of pre-processing and to evaluate the effect of different ranges of splitting training and test sets in the forecasting systems; (ii) reconstruct the decomposed signal through the weighted integration considering the no negative constraint theory; (iii) selection of the models for components through optimization techniques and considering the metaheuristics in order to obtain the best hyperparameters, such as the new approaches like [75–79]; (iv) developing an adaptive version of COA to define the number of coyotes and packs; and, (v) to perform cross-country comparisons using different forecasting models linking between models accuracy and economic aspects.

Author Contributions: M.H.D.M.R.: Conceptualization, Methodology, Formal analysis, Validation, Writing—Original Draft, Writing—Review & Editing; J.D.d.L.: Funding Acquisition, Writing—Review & Editing; S.F.S.: Writing—Review & Editing; A.N.: Writing—Review & Editing; V.C.M.: Conceptualization, Funding Acquisition, Writing—Review & Editing; L.d.S.C.: Conceptualization, Funding Acquisition, Writing—Review & Editing. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	Autoregressive Integrated Moving Average
CEEMD	Complementary Ensemble Empirical Mode Decomposition
COA	Coyote Optimization Algorithm
DM	Diebold-Mariano
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
ELM	Extreme Learning Machines
GBM	Gradient Boosting Machines
GP	Gaussian Process

IMF	Intrinsic Mode Function
IPEA	Institute of Applied Economics Research
LSTM	Long Short-Term Memory
LOOCV-TS	Leave-One-Out Cross-Validation Time Slice
MOWDT	Maximal Overlap Wavelet Discrete Transform
MWh	Mega-Watt Hour
OI	Ortogonal Index
OWA	Overall Weight Average
PACF	Partial Auto-Correlation Function
SVR	Support Vector Regression
sMAPE	Symmetric Mean Absolute Percentage Error
RMSE	Relative Mean Square Error
R^2	Coefficient of Determination
VMD	Variational Mode Decomposition

References

1. Abedinia, O.; Amjady, N.; Shafie-khah, M.; Catalão, J. Electricity price forecast using combinatorial neural network trained by a new stochastic search method. *Energy Convers. Manag.* **2015**, *105*, 642–654. [CrossRef]
2. Zhang, J.L.; Zhang, Y.J.; Li, D.Z.; Tan, Z.F.; Ji, J.F. Forecasting day-ahead electricity prices using a new integrated model. *Int. J. Electr. Power Energy Syst.* **2019**, *105*, 541–548. [CrossRef]
3. Muniain, P.; Ziel, F. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *Int. J. Forecast.* **2020**. [CrossRef]
4. Yang, Z.; Ce, L.; Lian, L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Appl. Energy* **2017**, *190*, 291–305. [CrossRef]
5. Maciejowska, K. Assessing the impact of renewable energy sources on the electricity price level and variability—A quantile regression approach. *Energy Econ.* **2020**, *85*, 104532. [CrossRef]
6. Corso, M.P.; Stefenon, S.F.; Couto, V.F.; Cabral, S.H.L.; Nied, A. Evaluation of methods for electric field calculation in transmission lines. *IEEE Lat. Am. Trans.* **2018**, *16*, 2970–2976. [CrossRef]
7. Aineto, D.; Iranzo-Sánchez, J.; Lemus-Zúñiga, L.G.; Onaindia, E.; Urchueguía, J.F. On the influence of renewable energy sources in electricity price forecasting in the Iberian market. *Energies* **2019**, *12*, 82. [CrossRef]
8. Brito, B.H.; Finardi, E.C.; Takigawa, F.Y. Unit-commitment via logarithmic aggregated convex combination in multi-unit hydro plants. *Electr. Power Syst. Res.* **2020**, *189*, 106784. [CrossRef]
9. Colonetti, B.; Finardi, E.C. Combining Lagrangian relaxation, benders decomposition, and the level bundle method in the stochastic hydrothermal unit-commitment problem. *Int. Trans. Electr. Energy Syst.* **2020**, 12514. [CrossRef]
10. Stefenon, S.F.; Freire, R.Z.; Coelho, L.S.; Meyer, L.H.; Grebogi, R.B.; Buratto, W.G.; Nied, A. Electrical insulator fault forecasting based on a wavelet neuro-fuzzy system. *Energies* **2020**, *13*, 484. [CrossRef]
11. Ribeiro, M.H.D.M.; Ribeiro, V.H.A.; Reynoso-Meza, G.; Coelho, L.S. Multi-objective ensemble model for short-term Price forecasting in corn price time series. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8. [CrossRef]
12. Ribeiro, M.H.D.M.; Coelho, L.S. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput.* **2020**, *86*. [CrossRef]
13. Ribeiro, M.H.D.M.; da Silva, R.G.; Mariani, V.C.; Coelho, L.S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos Solitons Fractals* **2020**, *135*. [CrossRef]
14. Zhang, W.; Qu, Z.; Zhang, K.; Mao, W.; Ma, Y.; Fan, X. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *136*, 439–451. [CrossRef]
15. Ribeiro, G.T; Sauer, J.G.; Fraccanabbia, N.; Mariani, V.C.; Coelho, L.S. Bayesian optimized echo state network applied to short-term load forecasting. *Energies* **2020**, *13*, 2390. [CrossRef]
16. Zhang, D.; Cai, C.; Chen, S.; Ling, L. An improved genetic algorithm for optimizing ensemble empirical mode decomposition method. *Syst. Sci. Control Eng.* **2019**, *7*, 53–63. [CrossRef]

17. Yang, W.; Wang, J.; Niu, T.; Du, P. A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting. *Appl. Energy* **2019**, *235*, 1205–1225. [CrossRef]

18. Zhang, J.; Tan, Z.; Wei, Y. An adaptive hybrid model for short term electricity price forecasting. *Appl. Energy* **2020**, *258*. [CrossRef]

19. Qiao, W.; Yang, Z. Forecast the electricity price of U.S. using a wavelet transform-based hybrid model. *Energy* **2020**, *193*, 116704. [CrossRef]

20. Zhang, X.; Wang, J.; Gao, Y. A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM. *Energy Econ.* **2019**, *81*, 899–913. [CrossRef]

21. Kasburg, C.; Stefenon, S. F. Deep learning for photovoltaic generation forecast in active solar trackers. *IEEE Lat. Am. Trans.* **2019**, *17*, 2013–2019. [CrossRef]

22. Yang, W.; Wang, J.; Niu, T.; Du, P. A novel system for multi-step electricity price forecasting for electricity market management. *Appl. Soft Comput.* **2020**, *88*. [CrossRef]

23. Zhou, S.; Zhou, L.; Mao, M.; Tai, H.; Wan, Y. An optimized heterogeneous structure LSTM network for electricity price forecasting. *IEEE Access* **2019**, *7*, 108161–108173. [CrossRef]

24. Khalid, R.; Javaid, N.; Al-zahrani, F.A.; Aurangzeb, K.; Qazi, E.U.H.; Ashfaq, T. Electricity load and price forecasting using Jaya-Long Short Term Memory (JLSTM) in smart grids. *Entropy* **2020**, *22*, 10. [CrossRef]

25. Alipour, M.; Aghaei, J.; Norouzi, M.; Niknam, T.; Hashemi, S.; Lehtonen, M. A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration. *Energy* **2020**, *205*, 118106. [CrossRef]

26. Kazemzadeh, M.R.; Amjadian, A.; Amraee, T. A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. *Energy* **2020**, *204*, 117948. [CrossRef]

27. Kaboli, S.H.A.; Fallahpour, A.; Selvaraj, J.; Rahim, N. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. *Energy* **2017**, *126*, 144–164. [CrossRef]

28. Heydari, A.; Majidi Nezhad, M.; Pirshayan, E.; Astiaso Garcia, D.; Keynia, F.; De Santoli, L. Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. *Appl. Energy* **2020**, *277*, 115503. [CrossRef]

29. Stefenon, S.F.; Silva, M.C.; Bertol, D.W.; Meyer, L.H.; Nied, A. Fault diagnosis of insulators from ultrasound detection using neural networks. *J. Intell. Fuzzy Syst.* **2019**, *37*, 6655–6664. [CrossRef]

30. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.S, L.; da Rocha, D.F.M.; Grebogi, R.B.; de Barros Ruano, A.E. Wavelet group method of data handling for fault prediction in electrical power insulators. *Int. J. Electr. Power Energy Syst.* **2020**, *123*, 106269. [CrossRef]

31. da Silva, R.G.; Ribeiro, M.H.D.M.; Mariani, V.C.; Coelho, L.S. Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. *Chaos Solitons Fractals* **2020**, *139*. [CrossRef]

32. Brito, B.; Finardi, E.; Takigawa, F. Mixed-integer nonseparable piecewise linear models for the hydropower production function in the unit commitment problem. *Electr. Power Syst. Res.* **2020**, *182*, 106234. [CrossRef]

33. Fredo, G.L.M.; Finardi, E.C.; de Matos, V.L. Assessing solution quality and computational performance in the long-term generation scheduling problem considering different hydro production function approaches. *Renew. Energy* **2019**, *131*, 45–54. [CrossRef]

34. Finardi, E.; Lobato, R.; de Matos, V.; Sagastizábal, C.; Tomasgard, A. Stochastic hydro-thermal unit commitment via multi-level scenario trees and bundle regularization. *Optim. Eng.* **2020**, *21*, 393–426. [CrossRef]

35. van Ackooij, W.; Finardi, E.C.; Ramalho, G.M. An exact solution method for the hydrothermal Unit commitment under wind power uncertainty with joint probability constraints. *IEEE Trans. Power Syst.* **2018**, *33*, 6487–6500. [CrossRef]

36. Gontijo, T.S.; Costa, M.A. Forecasting hierarchical time series in power generation. *Energies* **2020**, *13*, 3722. [CrossRef]

37. Wu, J.; Chen, Y.; Zhou, T.; Li, T. An adaptive hybrid learning paradigm integrating CEEMD, ARIMA and SBL for crude oil price forecasting. *Energies* **2019**, *12*, 1239. [CrossRef]

38. Niu, D.; Wang, K.; Sun, L.; Wu, J.; Xu, X. Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. *Appl. Soft Comput.* **2020**, *93*, 106389. [CrossRef]

39. Wu, J.; Zhou, T.; Li, T. Detecting epileptic seizures in EEG signals with complementary ensemble empirical mode decomposition and extreme gradient boosting. *Entropy* **2020**, *22*, 140. [CrossRef]

40. Yeh, J.R.; Shieh, J.S.; Huang, N.E. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2010**, *02*, 135–156. [CrossRef]

41. Pierezan, J.; Coelho, L.S. Coyote optimization algorithm: A new metaheuristic for global optimization problems. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8. [CrossRef]

42. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]

43. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.

44. Rasmussen, C.E. Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, Tübingen, Germany, 2–14 February 2003, 4–16 August 2003, Revised Lectures*; Springer: Heidelberg, Germany, 2004; pp. 63–71.

45. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

46. Zhao, X.; Wei, H.; Li, C.; Zhang, K. A hybrid nonlinear forecasting strategy for short-term wind speed. *Energies* **2020**, *13*, 1596. [CrossRef]

47. Kim, Y.; Hur, J. An ensemble forecasting model of wind power outputs based on improved statistical approaches. *Energies* **2020**, *13*, 1071. [CrossRef]

48. Stefenon, S.F.; Grebogi, R.B.; Freire, R.Z.; Nied, A.; Meyer, L.H. Optimized ensemble extreme learning machine for classification of electrical insulators conditions. *IEEE Trans. Ind. Electron.* **2020**, *67*, 5170–5178. [CrossRef]

49. Percival, D.B.; Walden, A.T. The maximal overlap discrete wavelet transform. In *Wavelet Methods for Time Series Analysis*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 2000; Chapter 5, pp. 206–254. [CrossRef]

50. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley Series in Probability and Statistics; John Wiley & Sons: Hoboken, NJ, USA, 2015; p. 712.

51. Athanasopoulos, G.; Hyndman, R.J. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Online. Available online: https://otexts.com/fpp2/ (accessed on 24 July 2020).

52. Assimakopoulos, V.; Nikolopoulos, K. The theta model: A decomposition approach to forecasting. *Int. J. Forecast.* **2000**, *16*, 521–530. [CrossRef]

53. Fiorucci, J.A.; Pellegrini, T.R.; Louzada, F.; Petropoulos, F.; Koehler, A.B. Models for optimising the theta method and their relationship to state space models. *Int. J. Forecast.* **2016**, *32*, 1151–1161. [CrossRef]

54. de Souza, R.C.T.; de Macedo, C.A.; Coelho, L.S.; Pierezan, J.; Mariani, V.C. Binary coyote optimization algorithm for feature selection. *Pattern Recognit.* **2020**, *107*, 107470. [CrossRef]

55. Pierezan, J.; Maidl, G.; Yamao, E.M.; Coelho, L.S.; Mariani, V.C. Cultural coyote optimization algorithm applied to a heavy duty gas turbine operation. *Energy Convers. Manag.* **2019**, *199*. [CrossRef]

56. Li, Z.; Cao, Y.; Dai, L.V.; Yang, X.; Nguyen, T.T. Optimal power flow for transmission power networks using a novel metaheuristic algorithm. *Energies* **2019**, *12*, 4310. [CrossRef]

57. Nguyen, T.T.; Nguyen, T.T.; Nguyen, N.A.; Duong, T.L. A novel method based on coyote algorithm for simultaneous network reconfiguration and distribution generation placement. *Ain Shams Eng. J.* **2020**. [CrossRef]

58. Yuan, Z.; Wang, W.; Wang, H.; Yildizbasi, A. Developed coyote optimization algorithm and its application to optimal parameters estimation of PEMFC model. *Energy Rep.* **2020**, *6*, 1106–1117. [CrossRef]

59. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]

60. Li, T.; Zhou, Y.; Li, X.; Wu, J.; He, T. Forecasting daily crude oil prices using improved CEEMDAN and Ridge regression-based predictors. *Energies* **2019**, *12*, 3603. [CrossRef]

61. Huang, N.; Xing, E.; Cai, G.; Yu, Z.; Qi, B.; Lin, L. Short-term wind speed forecasting based on low redundancy feature selection. *Energies* **2018**, *11*, 1638. [CrossRef]

62. Zhu, S.; Wang, X.; Shi, N.; Lu, M. CEEMD-subset-OASVR-GRNN for ozone forecasting: Xiamen and Harbin as cases. *Atmos. Pollut. Res.* **2020**. [CrossRef]

63. Mariani, V.C.; Och, S.H.; Coelho, L.S.; Domingues, E. Pressure prediction of a spark ignition single cylinder engine using optimized extreme learning machine models. *Appl. Energy* **2019**, *249*, 204–221. [CrossRef]

64. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems 9*; Mozer, M.C., Jordan, M.I., Petsche, T., Eds.; MIT Press: Cambridge, MA, USA, 1997; pp. 155–161. Available online: https://papers.nips.cc/book/advances-in-neural-information-processing-systems-9-1996 (accessed on 24 July 2020).

65. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [CrossRef]

66. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.

67. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.

68. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw. Artic.* **2008**, *28*, 1–26. [CrossRef]

69. Fiorucci, J.A.; Louzada, F.; Yiqi, B. *forecTheta: Forecasting Time Series by Theta Models*; R package version 2.2; 2016. Available online: https://cran.r-project.org/web/packages/forecTheta/forecTheta.pdf (accessed on 1 August 2020).

70. Hyndman, R.; Athanasopoulos, G.; Bergmeir, C.; Caceres, G.; Chhay, L.; O'Hara-Wild, M.; Petropoulos, F.; Razbash, S.; Wang, E.; Yasmeen, F. *Forecast: Forecasting Functions for Time Series and Linear Models*; R package version 8.12; 2020. Available online: http://pkg.robjhyndman.com/forecast (accessed on 1 August 2020).

71. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *26*, 1–22.

72. Zefrehi, H.G.; Altınçay, H. Imbalance learning using heterogeneous ensembles. *Expert Syst. Appl.* **2020**, *142*. [CrossRef]

73. Hao, Y.; Tian, C.; Wu, C. Modelling of carbon price in two real carbon trading markets. *J. Clean. Prod.* **2020**, *244*. [CrossRef]

74. Agrawal, R.K.; Muchahary, F.; Tripathi, M.M. Ensemble of relevance vector machines and boosted trees for electricity price forecasting. *Appl. Energy* **2019**, *250*, 54–548. [CrossRef]

75. de Vasconcelos Segundo, E.H.; Mariani, V.C.; Coelho, L.C. Design of heat exchangers using Falcon Optimization Algorithm. *Appl. Therm. Eng.* **2019**, *156*, 119–144. [CrossRef]

76. de Vasconcelos Segundo, E.H.; Mariani, V.C.; Coelho, L.C. Metaheuristic inspired on owls behavior applied to heat exchangers design. *Therm. Sci. Eng. Prog.* **2019**, *14*. [CrossRef]

77. Klein, C.E.; Mariani, V.; Coelho, L.S. Cheetah Based Optimization Algorithm: A Novel Swarm Intelligence Paradigm. In Proceedings of the 26th European Symposium on Artificial Neural Networks, ESANN, Bruges, Belgium, 25–27 April 2018; pp. 685–690.

78. Askarzadeh, A.; Coelho, L.C.; Klein, C.E.; Mariani, V.C. A population-based simulated annealing algorithm for global optimization. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 004626–004633. [CrossRef]

79. Galuzio, P.P.; de Vasconcelos Segundo, E.H.; dos Coelho, L.C.; Mariani, V.C. MOBOpt—Multi-objective Bayesian optimization. *SoftwareX* **2020**, *12*, 100520. [CrossRef]